

# Kazi Samin Mubasshir

Lawson Computer Science Building, 305 N University St, West Lafayette, IN 47907

✉ kmubassh@purdue.edu | 🌐 www.ksmubasshir.com

## Education

### Purdue University

West Lafayette, IN

Ph.D. in Computer Science

Fall 2022 - Spring 2027

- **Advisor:** Dr. Elisa Bertino
- **Research Topic:** Network Security, Network Traffic Classification (NTC), Intrusion Detection Systems (IDS)
- **Courses:** Security Analytics, Natural Language Processing, Network Security, Information Security, Distributed Systems, Data Communication and Computer Networks, Algorithm Design, Analysis and Implementations

### Bangladesh University of Engineering and Technology

Dhaka, Bangladesh

B.Sc. in Computer Science and Engineering

Feb 2016 - Feb 2021

- **Thesis:** Aligner Ensembling, Batch Filtering, and New Datasets for Bengali-English Machine Translation
- **Advisor:** Dr. Rifat Shahriyar
- Graduated with Distinction

## Work Experience

### Amazon Web Services (AWS)

Boston, Massachusetts

Applied Scientist Intern

Summer [2024, 2025]

- External Security Services (ESS)
- Security Analytics and AI Research (SAAR)

### Amazon

West Lafayette, IN

Campus Brand Ambassador (CBA)

Mar 2025 - Present

- Purdue University

### Purdue University

West Lafayette, IN

Graduate Teaching Assistant

Spring '25, Fall '25

- CS180: Problem Solving And Object-Oriented Programming (Spring '25, Fall '25)

### Advanced Chemical Industries (ACI) Ltd.

Dhaka, Bangladesh

Machine Learning Engineer

Sept 2020 - Aug 2022

- Management Information Systems (MIS)

## Research Interests

Network Security, Cloud Security, Intrusion Detection Systems, Network Traffic Classification, Machine Learning

## Research Positions

### Purdue University

West Lafayette, IN

Graduate Research Assistant

Fall 2022 - Fall 2024

- Cyber Space Security Lab (Cyber2Slab, PI: Dr. Elisa Bertino)

### Bangladesh University of Engineering and Technology

Dhaka, Bangladesh

Undergraduate Research Assistant

Mar 2019 - Feb 2021

- BUET CSE NLP Group (PI: Dr. Rifat Shahriyar)

## Publications

### CONFERENCE PROCEEDINGS

- **Kazi Samin Mubasshir\***, Imtiaz Karim\*, Elisa Bertino, **Gotta Detect 'Em All: Fake Base Station and Multi-Step Attack Detection in Cellular Networks**, The 34th USENIX Security Symposium (Usenix Sec), 2025, (\*joint first authors).
- Imtiaz Karim, **Kazi Samin Mubasshir**, Mirza Masfiquur Rahman, Elisa Bertino, **SPEC5G: A Dataset for 5G Cellular Network Protocol Analysis**, IJCNLP-AAACL 2023: The 13th International Joint Conference on Natural

- Karim, Imtiaz, Hyunwoo Lee, Hassan Asghar, **Kazi Samin Mubasshir**, Seulgi Han, Mashroor Hasan Bhuiyan, and Elisa Bertino. **"VWAttacker: A Systematic Security Testing Framework for Voice over WiFi User Equipments."** arXiv preprint arXiv:2508.01469 (2025).
- Abhik Bhattacharjee, Tahmid Hasan, **Kazi Samin**, Md Saiful Islam, M. Sohel Rahman, Anindya Iqbal, Rifat Shahriyar, **BanglaBERT: Combating Embedding Barrier for Low-Resource Language Understanding**, Findings of the Association for Computational Linguistics: NAACL 2022
- Tahmid Hasan, Abhik Bhattacharjee, Md Saiful Islam, **Kazi Samin**, Yuan-Fang Li, Yong-Bin Kang, M. Sohel Rahman, Rifat Shahriyar, **XL-Sum: Large-Scale Multilingual Abstractive Summarization for 44 Languages**, Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021
- Tahmid Hasan, Abhik Bhattacharjee, **Kazi Samin**, Masum Hasan, Madhusudan Basak, M. Sohel Rahman and Rifat Shahriyar, **Not Low-Resource Anymore: Aligner Ensembling, Batch Filtering, and New Datasets for Bengali-English Machine Translation**, Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)

## Research Experience

---

### SPEC5G, FBSDetector

Purdue University

Fall 2022 - Spring 2027

- **SPEC5G:** SPEC5G is the first-ever public 5G dataset for NLP research. By leveraging large-scale pre-trained language models that have achieved state-of-the-art results on NLP tasks, we use this dataset for security-related text classification and summarization.
- **FBSDetector:** FBSDetector is an effective and efficient detection solution that can reliably detect FBSes and multi-step attacks from Layer-3 network traces using machine learning at the user equipment (UE) side. To develop FBSDetector, we create the first-ever high-quality and large-scale dataset capturing the network traces in different real-world cellular network scenarios. Our novel machine-learning models, specially designed to detect FBSes and multi-step attacks can effectively detect FBSes with high accuracy and low false positive rates.

### Undergraduate Thesis

Bangladesh University of

Engineering and Technology

Mar 2019 - Feb 2021

### Not Low Resource Anymore, XLSum, BanglaBERT

- **Not Low-Resource Anymore:** We built a customized sentence segmenter for Bengali and proposed two novel methods for parallel corpus creation on low-resource setups: aligner ensembling and batch filtering. With the segmenter and the two methods combined, we compile a high-quality Bengali-English parallel corpus that was unavailable before.
- **XL-Sum:** XL-Sum is a comprehensive and diverse dataset comprising annotated article-summary pairs. To extract them, we used carefully designed heuristics. As human and intrinsic evaluation indicates, it is highly abstractive, concise, and of high quality.
- **BanglaBERT:** BanglaBERT is a BERT-based Natural Language Understanding (NLU) model pretrained in Bangla. We introduce two downstream task datasets on natural language inference and question answering, and benchmark four diverse NLU tasks covering text classification, sequence labeling, and span prediction. In the process, we bring them under the first-ever Bangla Language Understanding Benchmark (BLUB). BanglaBERT achieves state-of-the-art results outperforming multilingual and monolingual models.

## Academic Services

---

- **Graduate Student Mentor:**
  - Computer Science Graduate Student Board (CSGSB) Mentor-Mentee Program 2024, 2025
  - Mentored six Purdue CS Graduate Students
- **Conference Sub Reviewer:**
  - IEEE Symposium on Security and Privacy (IEEE S&P) 2025, 2026
  - ACM Conference on Computer and Communications Security (CCS) 2025
  - USENIX Security Symposium 2026
  - Network and Distributed System Security (NDSS) 2024, 2025, 2026
  - International Symposium on Research in Attacks, Intrusions and Defenses (RAID) 2023, 2024
  - International Conference on Distributed Computing Systems (ICDCS) 2023
  - ACM Conference on Data and Application Security and Privacy (CODASPY) 2023
- **Journal Sub Reviewer:**
  - ACM Transactions on Privacy and Security (TOPS)
- **Artifact Evaluation Committee (AEC) Member:**

- ACM Conference on Computer and Communications Security (CCS) 2024, 2025
- USENIX Security Symposium 2023
- 17th IEEE Workshop on Offensive Technologies (WOOT), co-located with IEEE S&P 2023

## Media Coverage and Software Artifacts

---

- **SPEC5G**: GitHub Repo: <https://github.com/SysNetS/SPEC5G>
- **FBS-Detector**: GitHub Repo: <https://github.com/SysNetS/fbsdetect>  
FBSDetector in News: Platforms for Advanced Wireless Research, Purdue Computer Science.

## Talks and Presentations

---

- Gotta Detect 'Em All: Fake Base Station and Multi-Step Attack Detection in Cellular Networks, 34th USENIX Security Symposium, Seattle, Washington, August 2025.
- Detecting Fake Base Stations in Cellular Networks, AWS-SAAR Tech Talks, Amazon, Boston, July, 2024.
- Fake Base Station Detection using PAWR Platforms, Young Gladiators-2023, Northeastern University, Boston, June, 2023.
- Fake Base Station Detection - Summary of research project conducted on the POWDER platform, MERIF-2023, Boston University, Boston, May, 2023.

## Awards and Achievements

---

- 2025 **Travel Award**, Purdue Graduate Student Government (PGSG)  
2025 **Graduate Travel Award**, Purdue College of Science (CoS)  
2025 **Student Grant**, USENIX Security Symposium  
2022 **Robi Datathon 2.0**, Runners-up  
2022 **AI for Bangla 1.0**, Second runners-up  
2017-20 **Dean's Listed**, Four consecutive years of the undergraduate studies

## Skills

---

- **Programming Languages**: C/C++, Java, Python, Julia, PHP, Javascript, 8086 Assembly Language, SQL
- **Network Simulators**: Cisco Packet Tracer, Wireshark, NS2, srsRAN, Open5GS
- **Web Frameworks**: Laravel, Django, .NET

## Reference

---

- **Dr. Elisa Bertino** (Ph.D. Advisor)  
Samuel D. Conte Professor of Computer Science  
Purdue University
- **Dr. Rifat Shahriyar** (Undergraduate Advisor)  
Professor  
Department of Computer Science and Engineering  
Bangladesh University of Engineering and Technology (BUET)
- **Dr. Imtiaz Karim** (Ph.D. Mentor)  
Assistant Professor  
Department of Computer Science  
Erik Jonsson School of Engineering and Computer Science  
The University of Texas at Dallas